Article:
Zhou, Rundong, Esfahani, Mohammad Nasr, Lu, Zhixiong et al. (2026) A data-driven framework for constructing representative operating cycles of high-horsepower tractors from field-road measurements. Measurement Science and Technology. 206109. ISSN: 0957-0233

https://doi.org/10.1088/1361-6501/ae6a12

A Data-Driven Framework for Constructing Representative Operating Cycles of High-Horsepower Tractors from Field-Road Measurements

Rundong Zhou[a,b], Mohammad Nasr Esfahani[c], Xiaoting Deng[a], Chao Su[b], Zhiru Zhou[b], Xiangfeng Kong[b], Gracie Harte[c], James Harrison[c], Zhixiong Lu[a]

(a. College of Engineering, Nanjing Agricultural University, Nanjing 210031, China

b. College of Locomotive Vehicle, Nanjing Vocational Institute of Railway Technology, Nanjing 210031, China

c. School of Physics, Engineering and Technology, University of York, YO10 5DD York, U.K)

1st Corresponding Author

Name: Zhixiong Lu

E-mail: 1274375811@qq.com

Postal address: 210031

Telephone number: 13951715780

2nd Corresponding Author

Name: Xiaoting Deng

F-mail: 2816344964@qq.com

Postal address: 210031

Telephone number: 15895992823

Abstract: accurate characterization of real-world operating behaviours of high-horsepower agricultural tractors is a fundamental challenge for subsequent modelling, analysis, and system design. The highly variable and multi-modal nature of field-road operations makes it difficult to construct operating cycles that faithfully represent measured dynamics while preserving statistical distributions and temporal correlation characteristics. This study proposes a data-driven framework for constructing representative operating cycles of high-horsepower tractors based on extensive field-road measurements. The framework integrates optimized K-means clustering guided by the Davies-Bouldin Index with vehicle specific power (VSP) per unit mileage as a physically informed descriptor. A reduced-dimensional feature space is established to characterize measured tractor dynamics, enabling the extraction of representative kinematic fragments from complex operational datasets. These fragments are then synthesized into a unified operating cycle that preserves the intrinsic statistical structure of the measured data. The representativeness of the constructed cycle is quantitatively evaluated using both characteristic parameter deviation and autocorrelation function (ACF) analysis. The average deviation of key kinematic features is 5.28%, while the RMSE between the measured and synthesized ACF curves is only 0.0312, indicating strong consistency in both statistical distribution and temporal continuity. To further assess practical applicability, Cruise-based energy consumption simulations are conducted under three distinct load conditions. The predicted fuel consumption differs from experimental measurements by less than 10%, confirming that the synthesized cycle effectively reproduces real operating behaviour within acceptable measurement

uncertainty. The proposed framework provides a systematic approach for measurement-based representation and validation of operating profiles of high-horsepower agricultural machinery, supporting further developments in modelling, evaluation, and control of agricultural vehicle systems.

Key words: data-driven framework, measurement-based modelling, operating cycle construction, high-horsepower tractors, field-road measurements

1. Introduction

Accurately representing real-world operating behaviour is a fundamental challenge in vehicle measurement and data-driven modelling. An operating cycle, commonly expressed as the temporal evolution of vehicle speed [1], serves not only as a descriptive summary of measured operation but also as a compact measurement representation of complex dynamic behaviour. The fidelity of such a representation directly affects the reliability of subsequent modelling, analysis and system evaluation. However, widely used standardized cycles, such as NEDC, FTP-75 and WLTC [2, 3], are primarily designed for certification purposes and often fail to capture the variability and statistical characteristics observed in localized real-world measurements. As a result, operating behaviours inferred from these cycles may deviate substantially from measured dynamics, limiting their effectiveness in measurement-based analysis and evaluation [4].

To improve the representativeness of standardized operating cycles, extensive efforts have been devoted to the development of measurement-based, data-driven cycle construction methodologies. These studies primarily aim to compensate for the inherent limitations of predefined test cycles by extracting representative operational patterns directly from measured vehicle data. In addressing the problem of operational pattern identification, numerous studies have adopted clustering and feature-space reduction techniques. Zhou et al [5] developed a UMAP-KNN-based framework for pure electric buses, enabling the extraction of representative speed-power characteristics from high-dimensional measurement data. Song et al [6] employed self-organizing map neural networks to incorporate road-gradient information into urban driving cycles, thereby improving measurement completeness under varying roadway conditions. For tracked vehicles, Hu et al [7] proposed a three-dimensional working-condition construction approach based on micro-motion segmentation and K-means clustering, effectively capturing multi-surface and multi-factor operational variability. Hull et al [8] further employed shape-based time-series clustering to construct representative operating cycles for auxiliary vehicles in South Africa, demonstrating the effectiveness of morphology-oriented pattern identification in measured operational data. Almachi et al [9] established representative urban driving cycles for Quito by combining geospatial analysis with K-means clustering, highlighting the role of spatially informed feature extraction in measurement-based cycle construction.

To preserve the temporal structure and stochastic characteristics of measured operation, alternative approaches have focused on micro-segmentation and probabilistic modelling. Li et al [10] constructed AMT tractor operating cycles using an improved short-stroke method combined with Markov-chain modelling, while Zhang et al [11] extended state-space representations to derive multidimensional tram driving cycles. Rueda and Daniel [12] further utilized multi-city NREL datasets to propose a generalizable micro-trip-based cycle construction

method for light-duty vehicles, emphasizing cross-region measurement consistency. Similar Markov-based and Monte Carlo-enhanced methodologies have been applied to electric vehicles, motorcycles and bus rapid transit systems, demonstrating improved agreement with measured operational statistics compared with conventional cycles [13-15].

With the availability of long-term naturalistic datasets, recent studies have further emphasized measurement representativeness and validation. Ashtari et al [16] constructed a plug-in electric vehicle driving cycle for Winnipeg using large-scale naturalistic data, illustrating the importance of long-term measurements in capturing real-world operational variability.Yang et al [17] proposed a big-data-driven framework for heavy-duty truck operating cycles and reported improved fuel-consumption prediction accuracy relative to existing standard cycles. Brady et al [18] developed an electric-vehicle driving cycle using six months of real-world data and demonstrated strong consistency with independent datasets. Bishop et al [19] quantified variability and accuracy trade-offs in realistic cycle construction through natural-driving experiments and stochastic modelling. Pouresmaeili et al [20] established a city-specific passenger-car driving cycle and calibrated emission estimates using on-board measurements, revealing substantial deviations from standardized European and American cycles.

Specific efforts have also been made toward reference working cycles for agricultural machinery. The development of reference working cycles for agricultural tractors utilized multi-year onboard measurements, clustering of operational states, and hidden Markov modelling to synthesize a representative tractor working cycle that better reflected real-world usage, particularly under low-load operating conditions. While this study highlights the feasibility of long-term measurement-based reference cycles for tractors, challenges remain in representing mixed field-road operation and capturing load-dependent variability using limited operational descriptors [21].

Overall, existing approaches for constructing vehicle operating cycles mainly rely on clustering-based methods, Markov chain models, and intelligent algorithm-assisted techniques [22-25]. Despite these advances, most existing approaches have been developed for on-road vehicles operating under relatively well-structured traffic conditions. Agricultural machinery, particularly high-horsepower tractors, remains largely underrepresented in measurement-oriented operating cycle research. In contrast to conventional road vehicles, tractors frequently operate in mixed field-road environments characterized by intermittent high-load events, frequent speed fluctuations and pronounced stochasticity. These features introduce substantial challenges for measurement-based representation, as conventional cycle-construction methods often struggle to preserve both the statistical structure and physical consistency of measured tractor dynamics. From a measurement perspective, the key challenge lies in constructing an operating cycle that can reliably represent the multidimensional characteristics of tractor operation while maintaining consistency with the underlying measured data. Existing methods

are not fully suited to capturing the combined effects of load variability, operational mode switching and non-stationary dynamics that are inherent to agricultural transport tasks. In recent studies, vehicle specific power (VSP) has been proven to be an effective physical descriptor for characterizing vehicle dynamic behaviours and energy consumption features, which provides a reliable basis for data-driven operating cycle construction for heavy-duty and agricultural equipment [26, 27]. Recent studies have also emphasized the importance of robust feature extraction and domain generalization for data-driven modelling under noisy and uncertain operating environments. Improving the adaptability of feature learning and clustering methods under varying field conditions is essential for ensuring the reliability of operating-cycle construction from measured tractor data. In related engineering systems, robust multi-scale feature learning and domain generalization strategies have been shown to improve data consistency and model reliability under noise disturbance and small-sample conditions [28-30], which provide useful methodological insights for operating-cycle construction under complex field-road environments.

To address this challenge, this study proposes a data-driven measurement framework for constructing representative operating cycles of high-horsepower tractors based on extensive field-road measurements. The proposed framework integrates optimized K-means clustering guided by the Davies-Bouldin Index (DBI) with vehicle specific power (VSP) per unit mileage as a physically informed descriptor. Measured kinematic data are first preprocessed and segmented, and a reduced-dimensional feature space is established to characterize tractor operating behaviour. Representative kinematic fragments are then identified through optimized clustering and synthesized into a unified operating cycle that preserves both the statistical distribution and the intrinsic temporal correlation of the measured data. To quantitatively evaluate the representativeness of the constructed cycle, characteristic parameter deviation and autocorrelation function (ACF) analysis are jointly employed, ensuring consistency in both kinematic features and short-term temporal continuity. Furthermore, a Cruise-based tractor model was used as an application example to assess fuel consumption under multiple load conditions. The results demonstrate that the synthesized cycle exhibits strong agreement with measured operational characteristics and effectively reproduces real-world operating behaviour, confirming its suitability as a reliable measurement-based representation of high-horsepower tractor operation.

2. Measurement data acquisition and kinematic segmentation

2.1 Measurement campaign design

To ensure that the constructed operating cycle reflected the representative transportation characteristics of high-horsepower tractors in southern hilly regions, the measurement campaign was designed to capture a wide range of road types and real-world operational scenarios. As

shown in Figure 1, the Jiangsu Modern Agricultural Machinery Technology Demonstration Park in Baima Town, Lishui District, Nanjing, and its surrounding areas were selected as the data collection site due to their diverse terrain and typical rural transportation features.

The test route encompassed multiple road surface types, including asphalt and concrete straight segments (⩾1 km), S-shaped curves, gravel roads, dirt roads, forest paths, and slopes with gradients ranging from 5% to 15%. In addition, real-world operational conditions such as village crossings, narrow road encounters, and mixed-traffic situations were incorporated to ensure comprehensive coverage of tractor transportation behaviours. This combination of varied terrains and authentic agricultural road environments provided a robust data set for subsequent operating cycle construction and analysis.

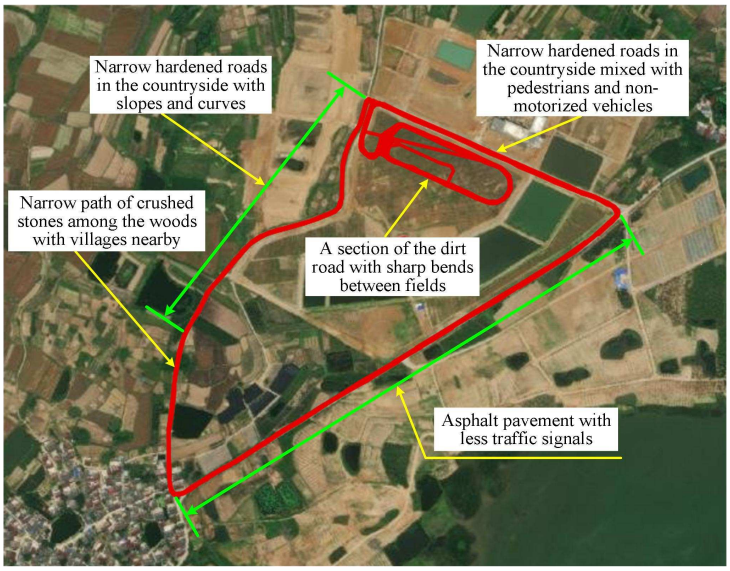

Figure 1. Data collection route encompassing representative road types relevant to tractor transportation operations

To ensure data representativeness while controlling testing duration, cost, and experimental constraints, the data collection strategy was designed based on typical tractor transportation scenarios, including agricultural supply delivery and product transfer. Three data collection methods were adopted.

(1) Route-based cycle testing. A predefined line-cycle test was conducted on major transportation routes to cover different road surfaces, including asphalt pavement, concrete pavement, gravel roads, and dirt roads. To reflect different operating load levels, tests were carried out under three loading conditions: 0%, 50%, and 100% of the trailer's rated capacity (i.e., empty-load, half-load, and full-load). The "multi-driver random rotation method" was used to reduce individual driving bias and improve data reliability.

(2) Hybrid tracking on typical segments. On representative rural road sections, such as village-to-field gravel roads and asphalt connector roads, tractors and light agricultural vehicles (e.g., tricycles and mini-trucks) were tracked using a tractor-agricultural vehicle hybrid tracking

approach. This method was used to establish comparable reference datasets under typical field-road transport operating cycles.

(3) Naturalistic driver operation. During specific time periods and on narrow road sections (approximately 3.5 m wide), data were collected using a naturalistic driving approach. Drivers operated the tractor according to their own driving habits and route preferences without additional intervention, allowing the acquisition of realistic operating behaviour data.

2.2 Measurement system and sensors

To ensure that the collected data accurately reflected the operational characteristics of tractors in the local test environment, factors such as trailer self-weight, road surface type, terrain slope, and installation feasibility were comprehensively considered when selecting the data acquisition equipment. Accordingly, the WTGAHRS1 inertial navigation sensor (Weite Co., Ltd.) was chosen for vehicle speed measurement. This sensor integrated high-precision GPS+BeiDou dual-navigation systems, providing a speed measurement accuracy better than 0.1 m s^{-1}, a positioning accuracy within 2.5 m, and an output rate adjustable from 0.2 to 200 Hz.

In addition to real-time recording, the sensor simultaneously measured auxiliary parameters including acceleration, angular velocity, attitude angle, magnetic field strength, barometric pressure, altitude, and geographic coordinates. These supplementary signals supported data screening and subsequent feature extraction. Considering that high-horsepower tractors under mixed field-road transportation operating cycles generally operate at relatively low speeds and exhibit smoother dynamic transitions than highway vehicles, the sampling frequency was set to 1 Hz through the upper-computer interface to capture key kinematic states such as acceleration, deceleration, idling, and steady driving. Meanwhile, this frequency effectively reduced data redundancy and computational burden during subsequent feature extraction and operating cycle construction, achieving a balance between data resolution and processing efficiency. The configuration of the sampling equipment system is shown in figure 2.

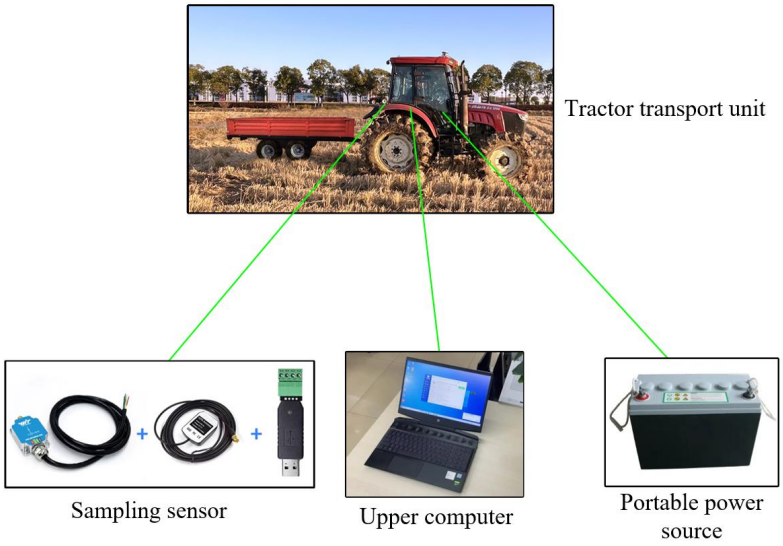

Figure 2. Experimental data acquisition system for tractor transportation tests, consisting of the tractor-trailer unit, a sampling sensor module (including speed, position, and dynamic signal collectors), an upper-computer terminal for real-time data monitoring and storage, and a portable power supply to ensure stable operation during field-road measurements.

2.3 Data preprocessing and kinematic segmentation

2.3.1 Preprocessing of measured data

Field-road operating data were collected during the typical autumn harvest season, which represented the busiest and most representative transportation period for high-horsepower tractors in southern hilly agricultural regions. During this period, tractors frequently performed mixed field-road transport tasks, including grain hauling, field transfer, and agricultural product transportation, resulting in high operational intensity and diverse load conditions. Compared with field tillage operations, transport operating conditions were less sensitive to seasonal meteorological variations, and their core kinematic characteristics, such as speed distribution, idle proportion, acceleration frequency, and speed fluctuation patterns, remained relatively stable across typical working periods. Therefore, the autumn harvest season provided sufficient statistical representativeness for constructing standard tractor transport operating cycles.

After approximately one month of continuous measurement, more than 100 sets of working-condition data with over 200,000 valid entries were obtained. During the data collection and transmission process, individual missing values, outliers, and abnormal speed fluctuations caused by road-surface interference were observed. Therefore, data preprocessing was required to eliminate errors and improve data reliability as follows:

(1) Handling missing values: Linear interpolation was performed when the number of consecutive missing values between two adjacent data points was fewer than 5. When the number of missing values exceeded 5, the corresponding data segment was removed to avoid

excessive interpolation errors and distortion of the original driving characteristics. Since a high-precision GPS+BeiDou dual-navigation sensor was mounted on the roof, no significant missing data occurred during the test.

(2) Handling excessively long idle time: If the idle time exceeded 180 s, it was considered that the tractor had entered a long-term parking state rather than normal operational idling. Therefore, only 180 s of idle time was retained [31].

(3) Removal of abnormal speed and acceleration data: The maximum travelling speed of the tractor does not exceed 30 km h^{-1}. Therefore, according to the national standard GB/T 33195-2016, data points with speeds exceeding 30 km/h, as well as outliers with accelerations greater than 2 m s^{-2} or decelerations less than -3.5 m s^{-2}, were removed to ensure the accuracy and reliability of the data.

A moving-average filter with a window size of 3 samples (equivalent to 3 s at a sampling frequency of 1 Hz) was applied to suppress high-frequency noise induced by road-surface irregularities. As shown in figure 3, the filtered speed profile preserves the trend of the original data while effectively removing high-frequency fluctuations, providing a more accurate representation of the tractor's actual operating speed.

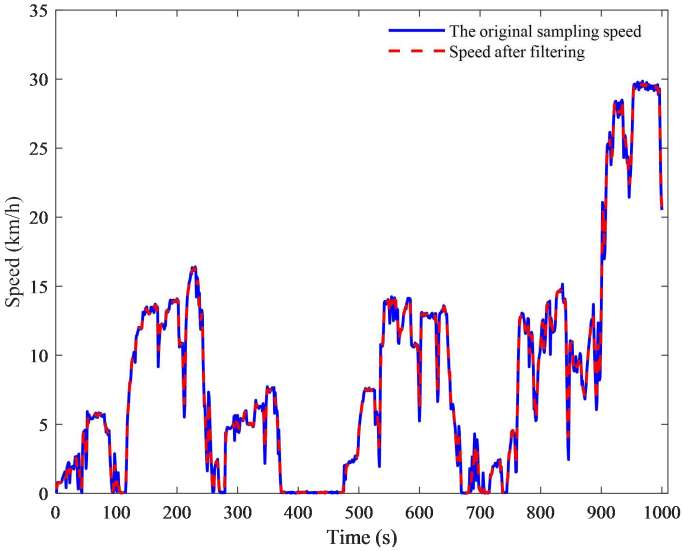

Figure 3. Comparison of raw speed data and moving-average filtered speed data

2.3.2 Kinematic segmentation

An idle segment is defined as the period from vehicle stop to the subsequent restart, while a kinematic segment refers to the driving process between two consecutive idle states. As shown in figure 4, a typical kinematic fragment contains four motion states, namely the acceleration state, the deceleration state, the uniform state and the idle state.Considering the operational efficiency and mechanical reliability of tractors, while taking into account the health of drivers and environmental protection requirements, the total duration of the operating conditions constructed in this study is 2,400 seconds.

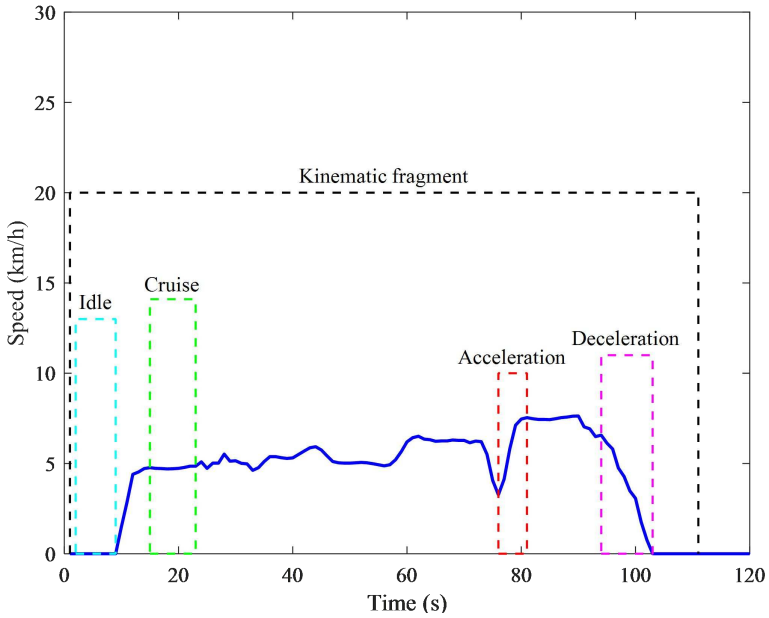

Figure 4. Kinematic fragment of tractor

In the construction of tractor transportation operating cycles, the speed-time curve is usually composed of several segments with specific kinematic characteristics. Due to differences in factors such as traffic environment, road conditions and time dimension, kinematic fragments show significant heterogeneity. The presence of short-term abnormal data and invalid fragments in the kinematic segment obtained from the initial division may lead to deviations in the calculation of characteristic parameters. To this end, a three-level kinematic fragment data screening mechanism for tractor was established:

(1) Integrity screening: When the interval between adjacent valid data points exceeds 180 seconds or the data loss rate exceeds 10%, it is determined as an incomplete fragment and eliminated.

(2) Duration screening: Eliminate abnormal segments with a duration of less than 30 seconds or more than 600 seconds.

(3) Motion state screening: According to the standard definition of GB/T 38146.2-2019, strictly distinguish: For the segments that do not meet the above definitions under acceleration conditions ($a \geq 0.15$ m s^{-2}), deceleration conditions ($a \leq -0.15$ m s^{-2}), uniform conditions ($/a/ < 0.15$ m s^{-2} and $v \geq 0.5$ km h^{-1}), and idle conditions ($a < 0.15$ m s^{-2} and $v < 0.5$ km h^{-1}), secondary filtering is carried out.

After three levels of screening, the invalid fragments in the original data set were effectively eliminated, and finally 349 effective kinematic fragments of tractor were obtained.

2.4 Selection and calculation of kinematic fragment characteristic parameters

2.4.1 Selection of characteristic parameters

The construction of tractor operating conditions is essentially a feature-driven pattern

recognition process. Representative kinematic fragments are selected from a classified fragment library based on optimized criteria, and a complete operating cycle is formed through time-series recombination. To mathematically describe each fragment and analyse its dynamic behaviour, a multidimensional characteristic parameter system must be established.

Analysis of the measured field-road transport data indicates that tractor operations exhibit distinct dynamic features: a start-stop frequency of 5-8 events per minute, a running-speed standard deviation exceeding 2.5 m s^{-1}, speed fluctuation ranges approaching 40% of the nominal speed, and peak acceleration-rate changes above 0.8 m s^{-2} during obstacle-avoidance maneuvers. Considering these characteristics: low-frequency, high-amplitude velocity fluctuations and frequent start-stop transitions, a total of 20 characteristic parameters selected in this study are shown in table 1 and divided into two major categories:

(1) Motion-state descriptive parameters.

These include average speed (with and without idle periods), average acceleration, maximum speed, speed standard deviation, maximum acceleration, maximum deceleration, and other indicators used to describe the dynamic behaviour within each fragment.

(2) Statistical distribution parameters.

These parameters characterize the temporal distribution of operating states, including idle-time proportion, speed-interval proportions, acceleration distribution, and related indicators. They capture the statistical patterns of fragment behaviour and serve as key metrics for validating the representativeness and effectiveness of the constructed tractor transportation operating cycle.

Table 1. Kinematic fragment parameters of tractor

Number	Characteristic parameters	Symbol	Unit
1	The duration of the kinematic fragment	T_p	s
2	Acceleration time	T_a	s
3	Deceleration time	T_d	s
4	Uniform time	T_c	s
5	Idle time	T_i	s
6	Average speed (including idle speed)	V_m	km h^{-1}
7	Average driving speed (excluding idle speed)	V_{mr}	km h^{-1}
8	Maximum speed	$V_{\max}$	km h^{-1}

Number	Characteristic parameters	Symbol	Unit
9	Standard deviation of velocity	V_{sd}	km h^{-1}
10	Average acceleration	A_m	m s^{-2}
11	Average deceleration	D_m	m s^{-2}
12	Standard deviation of acceleration	A_{sd}	m s^{-2}
13	Deceleration standard deviation	D_{sd}	m s^{-2}
14	Acceleration time ratio	P_a	%
15	Deceleration time ratio	P_d	%
16	Uniform time ratio	P_c	%
17	Idle time ratio	P_i	%
18	The proportion of 0-10 km h^{-1}	P_{0-10}	%
19	The proportion of 10-20 km h^{-1}	P_{10-20}	%
20	The proportion of 20-30 km h^{-1}	P_{20-30}	%

2.4.2 Calculation of characteristic parameters

Based on the selection results in Section 2.4.1, the 20 characteristic parameters selected in this study are defined as follows:

(1) The duration of the kinematic fragment T_p, acceleration time T_a, deceleration time T_d, uniform speed time T_c, idle time T_i.

$$T_p = \text{The duration of the kinematic fragment} \quad (1)$$

$$T_a = \text{The total duration of } A_i \geq 0.15 \text{ in this kinematic fragment} \quad (2)$$

$$T_d = \text{The total duration of } D_i \leq -0.15 \text{ in this kinematic fragment} \quad (3)$$

$$T_c = \text{The total duration of } abs(A_i) < 0.15 \text{ and } V_i \geq 0.5 \text{ in this kinematic fragment} \quad (4)$$

$$T_i = \text{The total duration of } abs(A_i) < 0.15 \text{ and } V_i \leq 0.5 \text{ in this kinematic fragment} \quad (5)$$

(2) Average speed (including idle speed) V_m, average driving speed (excluding idle speed) V_{mr}, maximum speed $V_{\max}$, speed standard deviation V_{sd}.

$$V_m = \frac{S}{T_p} \tag{6}$$

$$V_{mr} = \frac{S}{T - T_i} \tag{7}$$

$$V_{\max} = \max\{V_i \sim V_n\} \tag{8}$$

$$V_{sd} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(V_i - V_m)^2}, i = 1,2,\ldots n \tag{9}$$

Where, S represents the running distance of this kinematic segment, V_i indicates the velocity at the i-th second, in km/h, and n represents the total number of data points in this kinematic segment.

(3) Average acceleration A_m, average deceleration D_m, standard deviation of acceleration A_{sd}, standard deviation of deceleration D_{sd}.

$$A_m = \frac{sum\{A_i \mid A_i \geq 0.15, i = 1,2,\ldots n-1\}}{T_a} \tag{10}$$

$$D_m = \frac{sum\{A_i \mid A_i \leq -0.15, i = 1,2,\ldots n-1\}}{T_d} \tag{11}$$

$$A_{sd} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(A_i - A_m)}, i = 1,2,\ldots n \tag{12}$$

$$D_{sd} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(D_i - D_m)}, i = 1,2,\ldots n \tag{13}$$

Where, A_i and D_i represent the acceleration and deceleration of the segment in the i-th second, respectively, in m s^{-2}. The specific calculation formula is shown in formula (14).

$$A_i(D_i) = \begin{Bmatrix} 0 & i = 1 \\ \left(\frac{V_{i+1} - V_{i-1}}{2}\right)/3.6 & i = 2,3,...,n-1 \\ 0 & i = n \end{Bmatrix} \tag{14}$$

(4) Acceleration time ratio P_a, deceleration time ratio P_d, constant speed time ratio P_c, idle time ratio P_i.

$$P_a = \frac{T_a}{T_p} \tag{15}$$

$$P_d = \frac{T_d}{T_p} \tag{16}$$

$$P_c = \frac{T_c}{T_p} \tag{17}$$

$$P_i = \frac{T_i}{T_p} \tag{18}$$

(5) The proportion of 0-10 km h^{-1} P_{0-10}, that of 10-20 km h^{-1} P_{10-20}, and that of 20-30 km h^{-1} P_{20-30}.

$$P_{0-10} = \frac{T_{0-10}}{T_p} \tag{19}$$

$$P_{10-20} = \frac{T_{10-20}}{T_p} \tag{20}$$

$$P_{20-30} = \frac{T_{20-30}}{T_p} \tag{21}$$

Where, T_{0-10}, T_{10-20}, and T_{20-30} respectively represent the total number of velocities belonging to [0, 10), [10, 20), [20, 30] in this segment.

Based on the above formulas, the 349×20-dimensional kinematic fragment characteristic parameter value matrix Z of the tractor transportation condition was obtained, as shown in table 2.

Table 2. Kinematic Fragment Characteristic Parameter Matrix Z

Number	T_p	T_a	T_d	T_c	T_i	V_m	V_{mr}	V_{max}	…	P_{10-20}	P_{20-30}
1	105	10	14	77	5	2.86	2.92	5.77	…	0	0
2	165	22	32	100	12	9.97	10.61	16.32	…	0.66	0
3	102	7	9	78	9	4.91	5.32	7.63	…	0	0
4	298	19	33	144	103	5.94	9.01	14.09	…	0.35	0
5	317	59	69	182	8	14.57	14.90	29.64	…	0.31	0.3
6	182	24	28	115	16	10.70	11.66	16.95	…	0.62	0
7	567	58	69	317	124	7.39	9.43	23.63	…	0.39	0.02
8	66	6	10	32	19	6.00	8.20	10.77	…	0.35	0
…	…	…	…	…	…	…	…	…	…	…	…
348	639	61	60	367	152	15.07	19.29	30	…	0.18	0.44
349	338	38	43	69	189	6.97	14.77	29.6	…	0.09	0.18

Meanwhile, the overall operational characteristics of the processed test sample data were calculated, and the corresponding statistical pie chart. Figure 5 shows that the overall characteristics of the tractor's operating conditions are as follows: the proportion of idle speed and constant speed is relatively high, and the proportion of the medium and low speed range of

0-10 km h^{-1} is also large. This reflects the transportation operation characteristics of the tractor, such as long waiting time during the transfer operation and continuous driving at a stable speed.

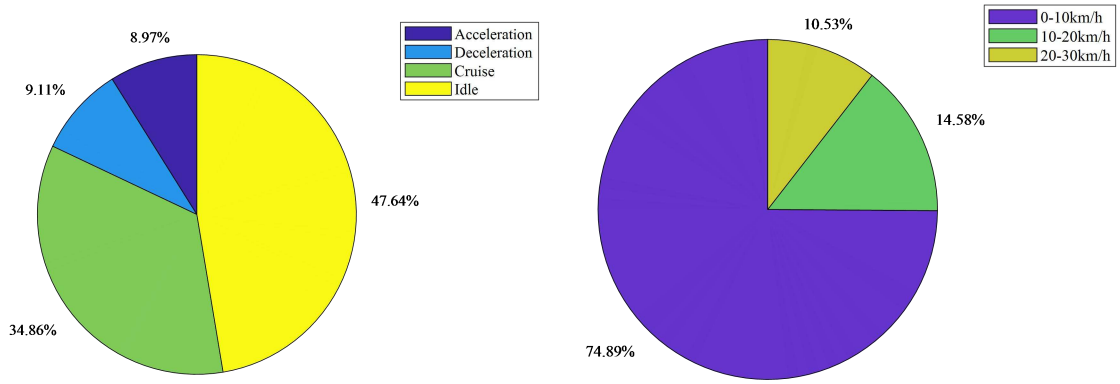

Figure 5. Overall velocity characteristic distribution of the test data

3. Measurement-based synthesis of operating cycles

As shown in figure 6, based on the heat map analysis of the Pearson correlation coefficient matrix, the interaction relationships among 20 characteristic parameters can be obtained. The diagonal elements of this matrix represent the autocorrelation of parameters (coefficient r =1), while the off-diagonal elements quantify the linear correlation strength between parameter pairs. A significant negative correlation was shown among them. Among time-domain parameters: the average speed (V_m) shows a strong positive correlation with the maximum travelling speed ($V_{\max}$) (r =0.9083). The statistical distribution parameters show modular characteristics: the proportion of idle time (P_i) and the standard deviation of speed (V_{sd}) form a negative correlation cluster (r =-0.4451); The dynamic characteristic parameters have independent information dimensions: the correlation coefficient between the standard deviation of deceleration (D_{sd}) and most other characteristic parameters is less than 0.35. Therefore, it is necessary to use principal component analysis to merge the characteristic parameters with large redundantly while retaining the independent parameters.

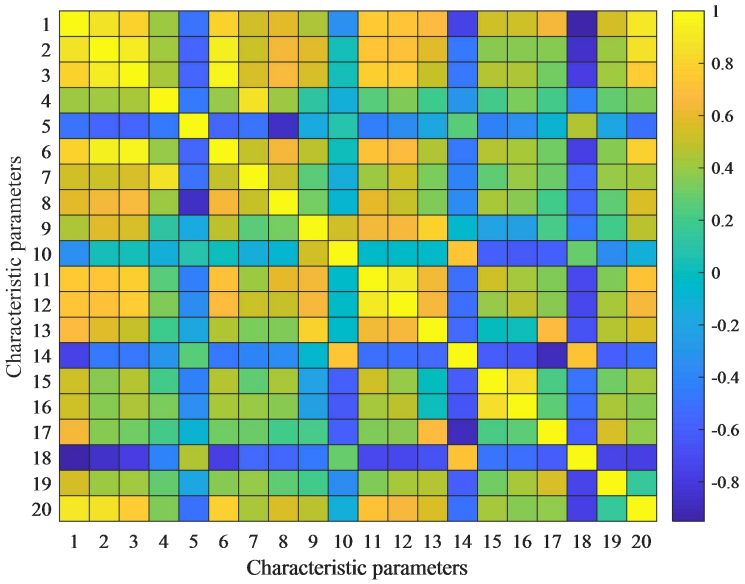

Figure 6. Heat map of the Pearson correlation coefficient matrix

3.1 Principal component analysis

3.1.1 PCA-based feature matrix construction

The sample contained 349 kinematic fragment data, and there are 20 characteristic parameters used to evaluate the sample data. PCA [32, 33] was employed to reduce redundancy among the 20 characteristic parameters while preserving the principal variance information of the original data set. Therefore, 349×20-dimensional matrix A can be obtained:

$$A=\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1,20} \\ a_{21} & a_{22} & \cdots & a_{2,20} \\ \vdots & \vdots & \ddots & \vdots \\ a_{349,1} & a_{349,2} & \cdots & a_{349,20} \end{bmatrix} \tag{22}$$

3.1.2 Principal component selection based on variance contribution

Based on PCA, the original 20-dimensional characteristic parameter matrix was reduced while preserving the dominant variance information of the datasets. The variance contribution rate and cumulative contribution rate of each principal component were then calculated, and the results are presented in table 3.

Table 3. Principal component variance contribution rate and cumulative contribution rate

Principal component	Principal component variance	Variance contribution rate	Cumulative contribution rate
1	10.326	51.629%	51.629%
2	3.013	15.064%	66.693%
3	2.017	10.087%	76.780%
4	1.394	6.969%	83.749%
5	0.869	4.347%	88.096%
6	0.765	3.822%	91.918%

7	0.693	3.467%	95.385%
8	0.287	1.436%	96.821%
9	0.167	0.835%	97.656%
10	0.117	0.586%	98.242%
11	0.106	0.532%	98.774%
12	0.096	0.481%	99.255%
13	0.077	0.385%	99.640%
14	0.038	0.191%	99.831%
15	0.016	0.083%	99.914%
16	0.012	0.061%	99.975%
17	4.844E-3	0.024%	99.999%
18	6.444E-5	3.222E-4	99.999E-4
19	1.012E-5	5.063E-5	99.999E-5
20	3.489E-32	1.721E-31	100%

It can be seen from table 3 that the first 7 principal components can explain 95.385% of the information of the original variable. To balance the information retention degree and the model complexity, they were selected as the effective feature subsets. Therefore, in order to further conduct cluster analysis, the principal component score matrix of the original sample data is shown in table 4.

Table 4. The principal component score matrix of the original sample

Number	PCA1	PCA2	PCA3	PCA4	PCA5	PCA6	PCA7
1	-1.460	-3.297	1.258	0.709	-0.408	-0.976	-1.772
2	2.988	-3.397	0.923	1.360	0.972	1.853	0.613
3	-1.008	-2.742	0.848	1.950	-1.451	-0.986	-1.538
4	1.435	-0.404	0.572	2.960	0.936	0.714	-0.609
5	6.679	-1.412	-0.121	0.1783	1.700	0.233	-2.016
6	3.335	-2.886	0.653	2.075	0.245	1.826	0.379
7	4.449	1.317	2.0127	1.489	1.672	2.357	-2.937
8	0.460	-2.918	-0.814	2.482	-0.400	0.598	0.318
…	…	…	…	…	…	…	…
348	5.983	3.201	2.675	-1.65	-0.006	0.128	-1.457
349	1.716	1.798	-1.075	-0.898	1.618	-0.143	-0.170

3.2 Clustering of kinematic fragments using DBI-optimised K-means

3.2.1 K-means clustering framework

In this study, the K-means clustering method [34] is used to study the test data of tractor transportation operating cycles. The core idea of the method is to assign each sample point to the category where the nearest mean (centroid) is located. The specific calculation steps are as follows:

(1) Initialize the clustering centres and randomly select n data points as the initial

clustering centres.

(2) For each sample $x_i = (x_{i1}, x_{i2}, ..., x_{im})$, calculate its distance to all centroids $c_n = (c_{n1}, c_{n2}, ..., c_{nm})$ (usually using Euclidean distance), and assign it to the cluster corresponding to the nearest centroid. The calculation formula is as follows:

$$d(x_i, c_n) = \sqrt{\sum_{j=1}^{m} (x_{ij} - c_{nj})^2} \tag{23}$$

Where, m represents the feature dimension;

(3) Update the cluster centre: For all samples in each cluster, calculate the mean of each feature dimension and take this mean as the new centroid.

$$c_n^{(new)} = \frac{1}{/S_n/} \sum_{x_i \in S_n} x_i \tag{24}$$

Where, $/S_n/$ represents the sample size of the i-th cluster;

(4) Iterate until convergence. Repeat steps 2 and 3 until one of the following conditions is met: the centroid change is less than the preset threshold (e.g., $\| c_n^{(new)} - c_n^{(old)} \| < \epsilon$, ϵ is the set threshold), and the maximum number of iterations is reached or the cluster allocation result no longer changes.

3.2.2 Determination of optimal cluster number using DBI

To determine the optimal number of clusters, this study uses the DBI to evaluate the clustering results. The clustering quality is assessed by calculating the ratio of the intra-cluster distance to the inter-cluster distance. The smaller the DBI value, the better the clustering effect [35, 36]. Its calculation method is as shown in equation (25):

$$\mathrm{DBI} = \frac{1}{\mathrm{k}} \sum_{i=1}^{\mathrm{k}} \max \frac{S(\mu_i) + S(\mu_j)}{M(\mu_i, \mu_j)}, \quad i \neq j \tag{25}$$

Where, k represents the number of clusters, and $S(\mu_i)$ is the average Euclidean distance from all points in cluster i to its clustering centre μ_i. $S(\mu_j)$ is the average Euclidean distance from all points in cluster j to its clustering centre μ_j. $M(\mu_i, \mu_j)$ represents the Euclidean distance between cluster centre μ_i and cluster centre μ_j.

As shown in figure 7, the DBI of the sample reaches its minimum value of 1.162 when k=2, and then increases with the increase of k, indicating that the clustering quality is optimal at this point. To further improve the robustness of cluster number determination, the Silhouette coefficient was additionally introduced as an auxiliary validation index. As shown in figure 8, the Silhouette analysis [37] also indicates that the optimal number of clusters is k=2, where the Silhouette coefficient reaches its maximum value of 0.493. The consistency of these two

independent validity indices confirms the robustness and reliability of the selected clustering result. Therefore, the kinematic fragments of the tractor are classified into two representative categories.

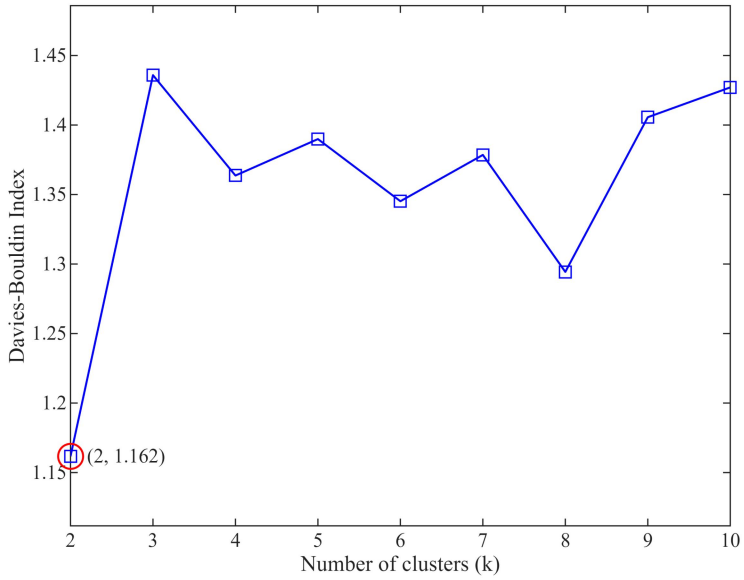

Figure 7. DBI versus number of clusters

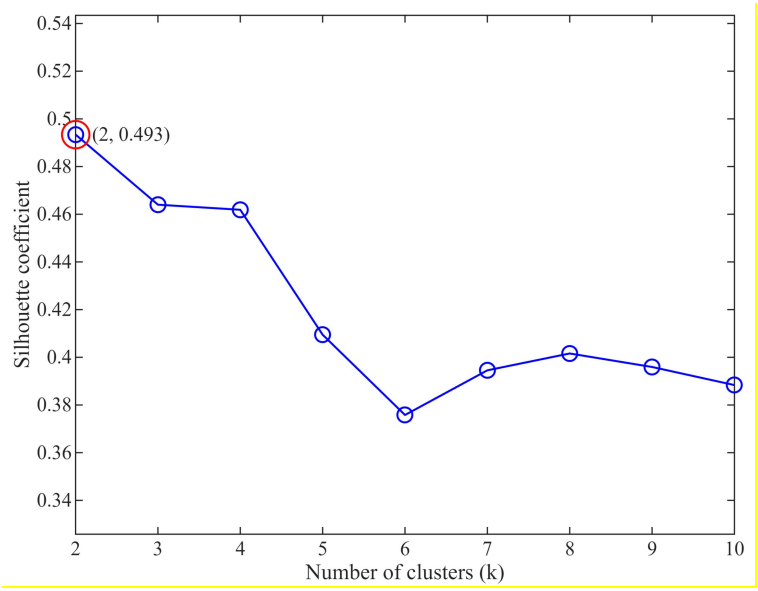

Figure 8. Silhouette coefficient versus number of clusters

The characteristic values of various fragment samples are shown in table 5. The sample numbers of the two types of fragments are 178 and 171 respectively, and the sampling distribution is relatively balanced. This indicates that both types of scenes are important in actual transportation, and the characteristic parameters between the categories are significantly different. The characteristics of category 1 are as follows: a relatively high average driving speed (14.563 km h^{-1}), a high proportion of constant speed (43%), strong short-term acceleration ability (20-30 km h^{-1} accounting for 21%), and a moderate proportion of idle speed (34%). These features are in line with the traffic conditions of county roads or rural roads and are

suitable for short-distance transportation scenarios in urban and rural areas. When in this category, the tractor can maintain stable driving on sections with fewer traffic signals. However, due to the interference of intersections and pedestrians, it still needs to stop intermittently, which makes the speed of 0-10 km h^{-1} account for 55.8%. The overall characteristics indicate medium and low-speed mixed transportation scenarios.

Category 2 is mainly characterized by: the main speed is low (0-10 km/h accounts for 94.1%), with occasional demands for medium and high speeds (5.6% for medium speed and 0.3% for high speed). The proportion of idle speed is relatively high (61.3%), and the acceleration and deceleration behaviours are similar (6.1% for acceleration and 6.5% for deceleration). These characteristics fully reflect the frequent start-stop and loading-unloading operation requirements of the field-road transportation operating cycles and the stability control requirements when towing heavy equipment, which precisely match the scenarios of internal field transportation and rural road transportation. Meanwhile, the average driving speed of category 2 (5.429 km h^{-1}) is slightly higher than the actual working conditions in the field (usually less than 5 km/h), but the proportion of low speed and idle speed can still effectively characterize the typical features of the transportation operation scenarios of high-horsepower tractor.

Table 5. The eigenvalues of various optimized fragment samples

Category	V_m	V_{mr}	P_a	P_d	P_c	P_i	•••	$P_{0–10}$	$P_{10–20}$	$P_{20–30}$	Number of fragments
1	9.928	14.563	0.116	0.114	0.430	0.340	…	0.558	0.232	0.210	178
2	2.342	5.429	0.061	0.065	0.261	0.613	…	0.941	0.056	0.003	171

3.3 Synthesis of representative operating cycle based on VSP-informed selection criteria

To achieve a representative measurement-based operating cycle, a synthesis strategy is proposed by jointly considering eigenvalue-based feature deviation and VSP per unit mileage. The overall procedure of operating-cycle synthesis is illustrated in figure 9. To balance measurement representativeness, computational efficiency and practical applicability, the total duration of the synthesized operating cycle is set to 2400 s. This duration is sufficient to capture the dominant operating modes observed in the field-road measurement data set while avoiding excessive redundancy in the reconstructed cycle.

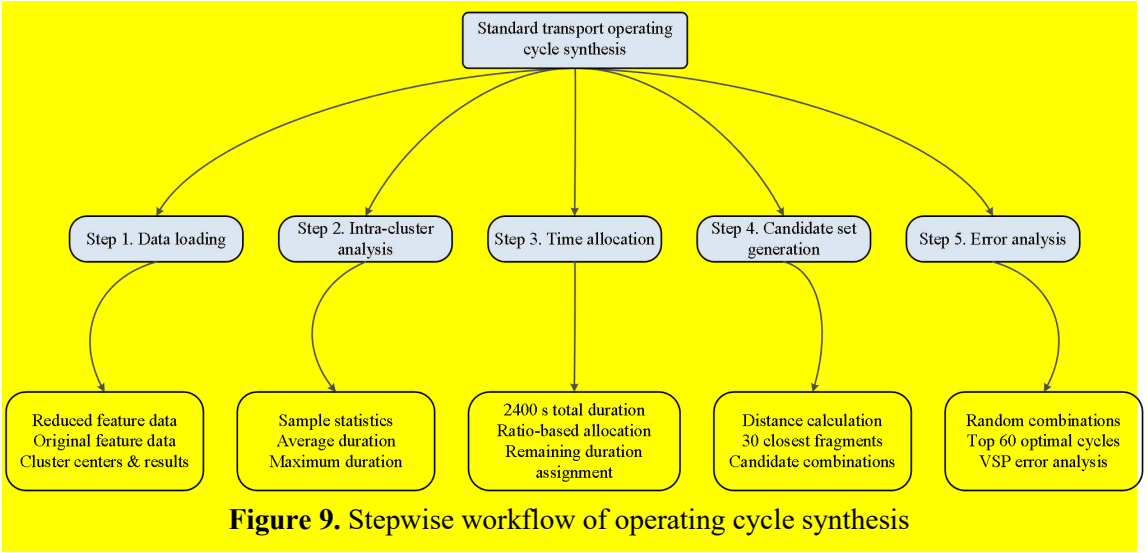

Figure 9. Stepwise workflow of operating cycle synthesis

The duration allocation for each category is derived from formula (26):

$$T_t = \frac{t_t}{T} \times 2400 \qquad (t = 1, 2) \tag{26}$$

Where, T_t represents the time occupied by the t-th type kinematic fragment in the final synthetic operating condition curve; t_t represents the total duration of all kinematic segments of Class t, and T indicates the total duration of all kinematic segments.

VSP [38] provides a more physically meaningful representation of vehicle operating load by jointly considering speed, acceleration, slope, and resistance factors. Its unit is kW t^{-1}. According to the national standard "Technical Conditions for the Operation Safety of Motor Vehicles" (GB7258-2017), the specific power of the tractor transport unit must be greater than or equal to 4.0 kW t^{-1}. The calculation formula for VSP is:

$$\mathrm{VSP} = v \times [a(1+\varepsilon) + g \times grade + C_R g] + \frac{\rho_a C_D A_s (v_w + v)^2 \times v}{2M} \tag{27}$$

Where, v represents the speed of the tractor, m s^{-1}, a is the acceleration of the tractor, m s^{-2}, ε is the mass factor, 0 for heavy-duty vehicles, g is the acceleration due to gravity, 9.8 m s^{-2}, $grade$ is the slope, 0.06, C_R is the rotational resistance coefficient of the tire, 0.03, ρ_a is the air density, 1.2 kg m^{-3}, C_D is the wind resistance coefficient, 0.7. A_s represents the windward area, 6.2 m^2. v_w represents the windward speed of the tractor, m s^{-1}. Since the maximum transport speed of the tractor in this study does not exceed 30 km/h and aerodynamic resistance contributes much less to the total tractive demand than rolling resistance, grade resistance, and inertial resistance, $v_w = 0$. M represents the total mass of the tractor and transport unit, 12,375 kg.

Based on formula (28), 60 candidate operating condition curves that meet the requirements of operating condition time length and the maximum VSP constraint are screened out.

$$\begin{cases} E = \sum_{n=1}^{20} \left(\dfrac{\left| \overline{F}_n^{syn} - \overline{F}_n^{real} \right|}{\overline{F}_n^{\text{real}}} \times 100 \right) \\ C_{60} \in \arg topE_{60} \\ s.t. \\ \sum_{i=1}^{N_1} T_{1i} + \sum_{j=1}^{N_2} T_{2j} = 2400 \pm 20 \\ VSP(C_{60})_{\max} \geq 4 \end{cases} \tag{28}$$

Where, E represents the sum of the average relative errors of all feature dimensions, $\overline{F}_n^{syn}$ is the average value of the synthetic working condition on the n-th feature, and $\overline{F}_n^{real}$ is the overall average value of the original data on the n-th feature. $\sum_{i=1}^{N_1} T_{1i}$ is the total time of the first type of kinematic segments (N_1) in the synthetic condition, and $\sum_{j=1}^{N_2} T_{2j}$ is the total time of the second type of kinematic segments (N_2) in the synthetic condition. The sum of the two satisfies the constraint of a total duration of 2400±20 s. C_{60} represents 60 candidate working conditions, and $\arg topE_{60}$ indicates the top 60 combinations with the smallest error in taking the eigenvalues.

In this study, VSP per unit mileage was selected as the physical descriptor because it provides a more comprehensive representation of tractor operating load than purely kinematic or torque-based indicators. Previous studies on heavy-duty vehicles and agricultural machinery have demonstrated that VSP-based descriptors are more effective than purely kinematic indicators for reflecting real driving load and energy consumption characteristics [26, 27].

Acceleration mainly describes local speed variation, but it cannot independently reflect the effects of road grade, rolling resistance, aerodynamic resistance, and vehicle mass. Torque-related indicators can directly characterize load demand; however, reliable torque measurements are difficult to obtain continuously during long-duration field-road tests without additional driveline instrumentation.

By contrast, VSP integrates vehicle speed, acceleration, road grade, rolling resistance, aerodynamic resistance, and total vehicle mass, thereby providing a physically interpretable measure of power demand under mixed field-road transportation operating cycles. In addition, normalizing VSP by mileage allows kinematic fragments with different durations and distances to be compared on a consistent basis.

Thus, this study calculates the VSP per mileage called P_L [kW (t·km)$^{-1}$] of the tractor from

514 the 60 candidate operating condition curves with the smallest feature deviation errors, as the
515 reference for the final operating condition selection, as shown in formula (29).

516
$$\begin{cases} P_L(c_n) = \dfrac{\sum_{i=1}^{N} VSP}{L} = \dfrac{\sum_{i=1}^{N} [v(a+0.882)+0.0002104v^3]}{L} \\ c_{best} = \underset{c_n \in C_{60}}{argmin} \left| P_L(c_n) - P_L(mean) \right| \end{cases} \tag{29}$$

517 Where, L represents the total mileage of the candidate working conditions, km. N
518 represents the number of data points for each candidate working condition. C_{best} represents the
519 final selected working condition; $P_L(mean)$ represents the VSP per mileage of the overall
520 sample data; and $P_L(c_n)$ represents the VSP per mileage of the candidate working conditions.

521 The final calculated VSP per mileage of the overall sample data was 893.0871[kW (t·km)$^{-1}$],
522 and the specific parameter statistics of the 60 candidate working conditions are shown in
523 table 6. Figure 10 shows that the absolute deviation between Candidate Cycle No. 14 and the
524 overall sample data was the smallest, with a value of 0.0140.

525 **Table 6.** Statistics of specific power per mileage under candidate working conditions

Candidate working conditions	Maximum VSP (kW t^{-1})	The sum of VSP (kW t^{-1})	Total mileage (km)	$P_L(c_n)$ [kW (t·km)$^{-1}$]	Absolute deviation [kW (t·km)$^{-1}$]
1	8.9029	5545.5829	6.2041	893.8578	0.7707
2	10.9778	3684.3602	4.1103	896.3726	3.2855
3	9.9732	4103.0857	4.5914	893.6459	0.5588
4	11.3015	3540.3505	3.9616	893.6668	0.5797
5	10.5915	3340.9823	3.7319	895.2497	2.1626
6	9.0792	3395.2378	3.8045	892.4268	0.6603
7	8.6534	4712.2987	5.2837	891.8558	1.2313
8	9.9615	2708.3171	3.0300	893.8340	0.7469
9	9.5190	5164.3565	5.7953	891.1284	1.9587
10	9.9615	3669.9984	4.1236	889.9986	3.0885
11	8.7290	3765.3498	4.2150	893.3214	0.2343
12	7.8478	3898.3363	4.3767	890.7022	2.3849
13	8.0569	3670.4435	4.1167	891.5985	1.4886
14	10.1289	4912.1454	5.5001	893.0871	0.0140
…	…	…	…	…	…
59	10.0880	3766.1043	4.2171	893.0555	0.0316
60	8.7518	3496.4774	3.9170	892.6417	0.4454

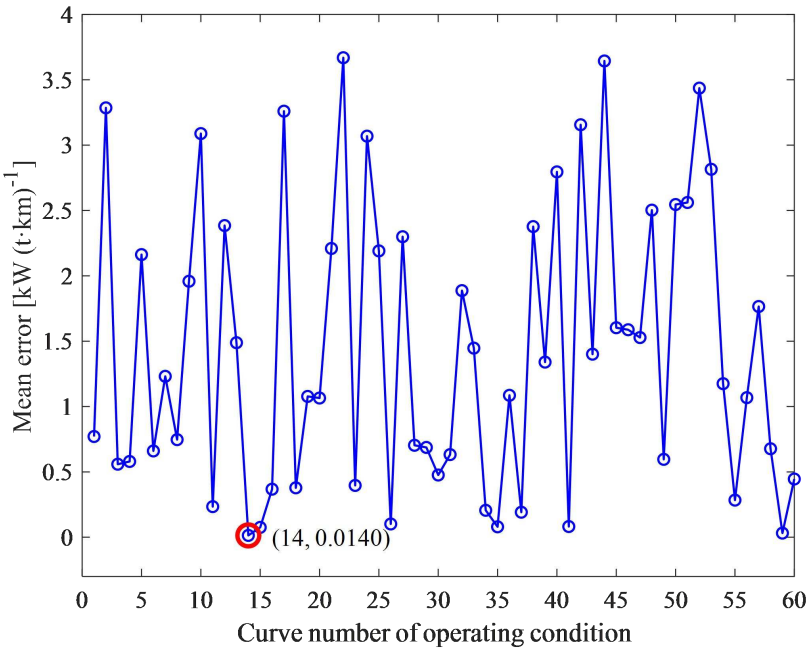

Figure 10. Absolute deviation of unit mileage specific power of candidate operating cycles from the overall sample mean

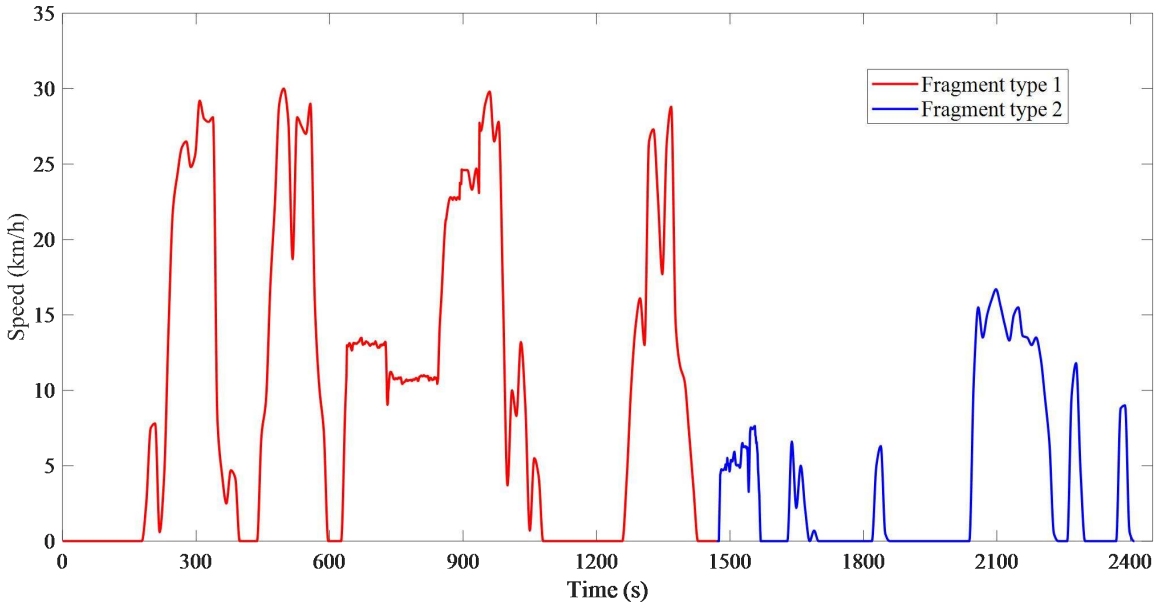

Figure 11. The constructed tractor's transportation working condition

The final synthesized operating cycle is shown in figure 11, with a total duration of 2,409 seconds. Among them, the first type of kinematic fragment accounts for 61.711%, and the second type of kinematic fragment accounts for 38.289%. The tractor speed exhibits significant variations during transportation operations. When the tractor is transporting within the field, the speed curve shows frequent start-stop and low-speed fluctuations (5-10 km h^{-1}). When the tractor is transporting on the road, the speed fluctuates periodically in the medium to high-speed range (20-30 km h^{-1}), corresponding to the deceleration-acceleration operation when avoiding obstacles or pedestrians, and occasionally there are brief medium-speed stable segments. When

the tractor is engaged in short-distance transportation between urban and rural areas, the overall speed is relatively high, but there are also sharp speed drops caused by sudden deceleration at intersections or pedestrian interference. These speed variations reflect the tractor's switching between different operation links during transportation, different road conditions (such as flat roads, slopes, curves, etc.), different load situations (such as empty, half-load, full load), and adjustments in driving strategies.

4. Validation of the constructed operating cycle

4.1 Error inspection of characteristics

The characteristic parameter error test of the constructed tractor transportation operation conditions was carried out. Twelve representative characteristic parameters, including average speed, average travelling speed, maximum speed, standard deviation of speed, average acceleration, and average deceleration, which are critical for evaluating the representativeness of the constructed operating cycle, are selected for error analysis. The results are shown in table 7. The characteristic parameters of the constructed tractor transportation cycle show good agreement with the measured data. The relative errors between the selected characteristic parameters and the overall sample data remain below 10%, and the overall average error is 5.28%. This indicates that the constructed tractor transportation operating cycle can accurately reflect the actual transportation situation in terms of characteristic parameters and have high practical value.

Table 7. The errors between the constructed working condition and the overall sample

Characteristic parameters	V_m	V_{mr}	$V_{\max}$	V_{sd}	A_m	D_m	A_{sd}	D_{sd}	P_a	P_d	P_c	P_i
Constructed working condition	6.813	10.59	30.000	6.418	0.267	-0.274	0.081	0.072	0.086	0.077	0.322	0.455
Overall sample	6.211	11.060	30.000	6.169	0.260	-0.257	0.077	0.069	0.081	0.083	0.349	0.476
Relative error	9.69%	5.27%	0	4.04%	2.69%	6.61%	5.19%	4.35%	6.17%	7.23%	7.74%	4.41%
Average error	5.28%											

4.2 Temporal correlation validation based on speed ACF

To further verify whether the synthesized driving cycle preserves the temporal dependency of the measured operating conditions, the speed autocorrelation function (ACF) was introduced for comparison. The ACF reflects the temporal continuity and dynamic persistence of speed

fluctuations, which is essential for evaluating the representativeness of micro-trip sequencing.

The ACF [39] at lag e is defined as:

$$R(e)=\frac{\sum_{t=1}^{N-e}\left(v_t-\overline{v}\right)\left(v_{t+e}-\overline{v}\right)}{\sum_{t=1}^{N}\left(v_t-\overline{v}\right)^2} \tag{30}$$

Where, $R(e)$ is the autocorrelation coefficient at lag e, $v(t)$ is the vehicle speed at time t, $\overline{v}$ is the mean speed, and N is the total number of samples.

The ACF describes the temporal dependency of vehicle speed. A slower decay indicates stronger continuity and smoother speed transitions, while a faster decay reflects more frequent fluctuations and transient changes.

Because the measured dataset is much longer than the synthesized cycle, the measured speed data were divided into multiple sliding windows with the same duration as the synthesized cycle. To ensure objective comparison, windows with the closest speed-related characteristics (mean speed, speed variation, idle ratio, acceleration/deceleration ratio, and fluctuation intensity) were selected as reference samples, and their averaged ACF was used for comparison.

The lag range of 0-40 s was selected because it captures the typical short-term speed transitions of tractor transport operations, such as stop-start, deceleration, and re-acceleration. Beyond this range, the ACF gradually approaches zero and is more affected by random disturbances, making it less representative for temporal-correlation evaluation.

To quantitatively evaluate the similarity between the two ACF curves, the root mean square error (RMSE) was calculated as:

$$\mathrm{RMSE}=\sqrt{\frac{1}{\mathrm{M}}\sum_{i=1}^{\mathrm{M}}\left(R_m(i)-R_s(i)\right)^2} \tag{31}$$

Where, $R_m(i)$ and $R_s(i)$ represent the ACF values of the measured samples and the synthesized cycle, respectively, and M is the number of lag points.

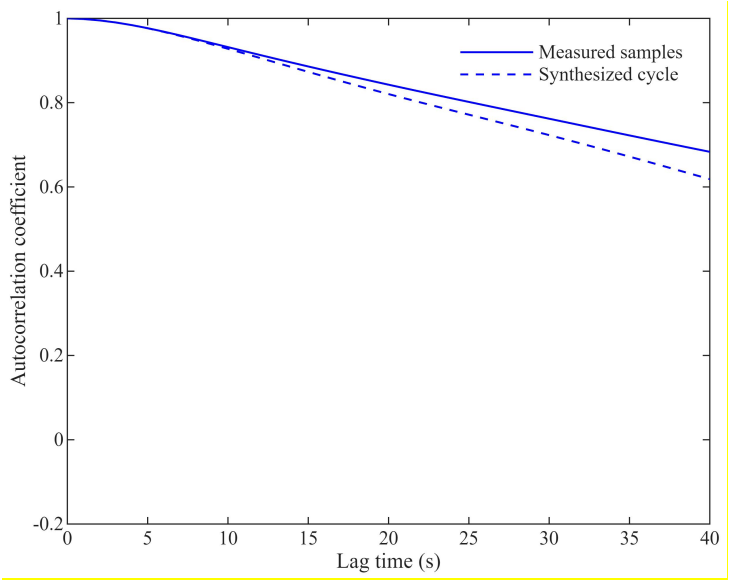

Figure 12. Comparison of ACF curves between measured samples and synthesized cycle

As shown in figure 12, the ACF curves of the synthesized cycle and the measured samples exhibit highly similar attenuation trends within the lag range of 0-40 s. Both curves decrease gradually with increasing lag time, indicating that the influence of the current speed state on future speed states weakens progressively. The RMSE between the two ACF curves was only 0.0312, which demonstrates that the synthesized cycle successfully preserves the short-term temporal dependence of the measured operating conditions.

This result further confirms that the proposed DBI-Kmeans and VSP-based driving cycle construction method not only reproduces the macroscopic statistical distribution characteristics, but also maintains the intrinsic microscopic temporal continuity of actual tractor transport operations.

4.3 Application-based validation using energy consumption simulation

The main component parameters of the test tractor transport unit are shown in table 8. To verify the fuel consumption characteristics of the synthesized operating cycle, a tractor simulation model was built based on Cruise software, and the mechanical and signal connections are completed. As shown in figure 13, the overall model includes the tractor body, the two-axle trailer, the braking system, the transmission system and the suspension system, etc. In the simulation, the trailer loads are respectively set to empty, half-load and full-load states. To be closer to the actual operation data, the simulation step size is set to 0.01 seconds. As shown in figure 14, the simulation operating cycle is set to the constructed tractor condition road spectrum.

Table 8. Parameters of tractor's transport unit

Parameter	Technical specifications	Explanation
Tractor model	Wode Aolong DX1204	Four-wheel drive (4WD) agricultural tractor
Engine	Inline four-cylinder	Displacement: 4.5 L

Parameter	Technical specifications	Explanation
	turbocharged diesel engine	
Rated power (kW)	88.2	Rated speed: 2300 r min^{-1}
Fuel tank capacity (L)	150	The battery life is approximately 8 to 10 hours (Medium-load operation)
Overall dimension (mm)	4552×2078×2987	A cab and standard tire configuration are included
Wheelbase (mm)	2330	The minimum turning radius is 5500 mm
Minimum usage weight (kg)	4375	Counterweight blocks and standard configuration are included
Maximum traction force (kN)	≥38	The maximum load capacity of the matching trailer is 8 to 10 tons (including its own weight)
Maximum lifting force (kN)	≥25	Compatible with hydraulic farm tools
Trailer model	Woli 12101	Self-unloading trailer
Vehicle body dimension (mm)	3800×2000×500	Double-axis type

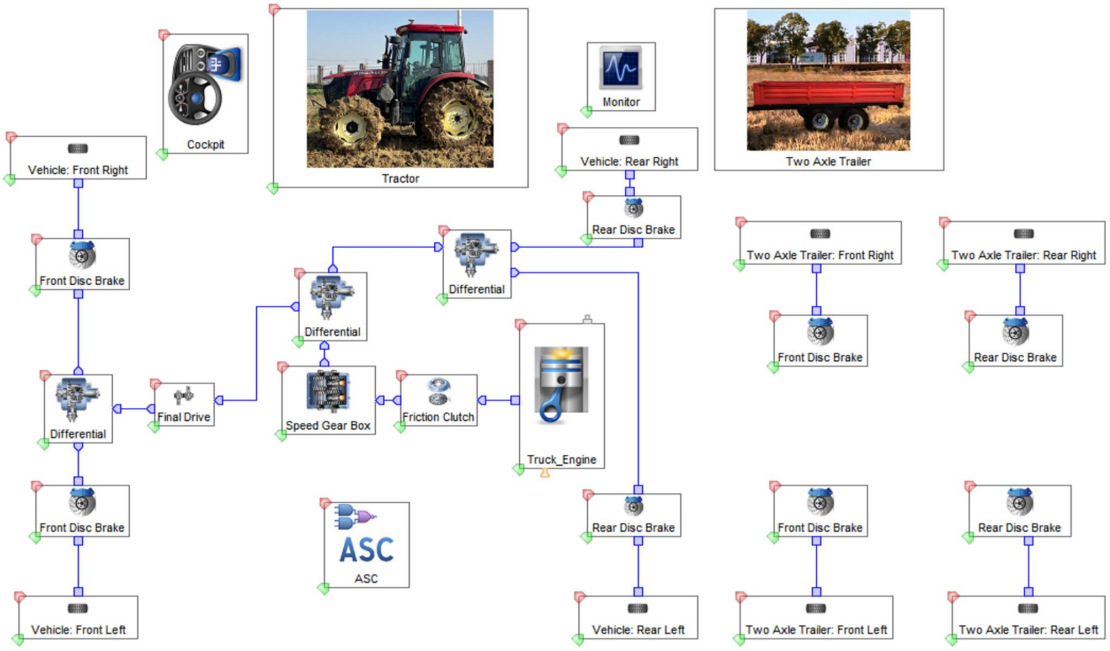

Figure 13. Cruise simulation model of tractor transport unit

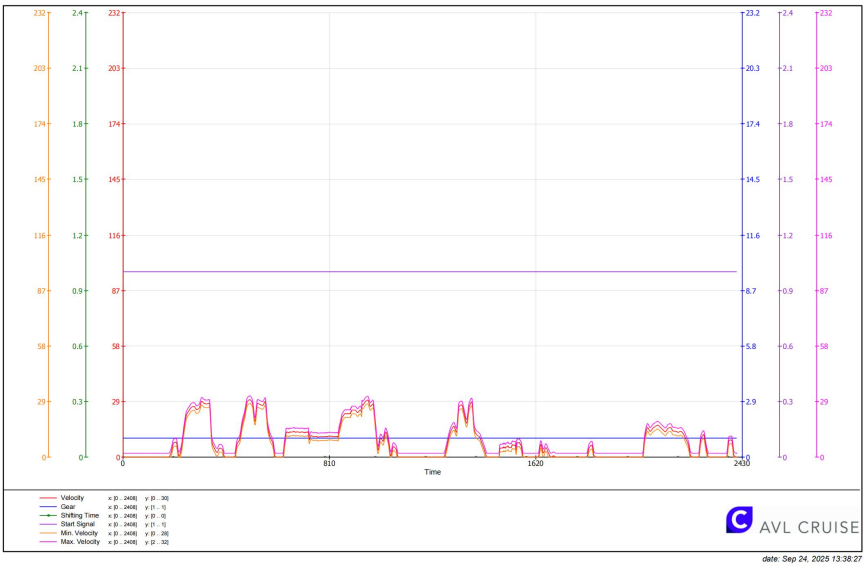

Figure 14. The setting of the Cruise simulation operating cycle

The tractor travelled a total distance of 5.286 kilometres. As shown in figure 15, when the tractor is empty, half, and full-load respectively, the real-time operating speed of the tractor remained within the upper and lower deviation bounds, and the speed tracking performance showed good agreement. Figure 16 shows that the trend of acceleration is consistent with that of velocity. Under the empty-load condition, due to the tractor's light weight and small inertia, its acceleration response is more sensitive compared to when it is half-load or full-load. This accurately reflects the transportation operation characteristics of tractors. Meanwhile, the fuel consumption characteristic data of the tractor under three working conditions are obtained and compared respectively with the test data. As shown in table 9, the fuel consumption per 100 kilometres of the tractor under empty-load, half-load and full-load conditions has an error of less than 10% compared with the actual tested fuel consumption of the tractor, indicating that the constructed tractor transportation operating cycle truly reflects the actual transportation operation characteristics of high-horsepower tractor.

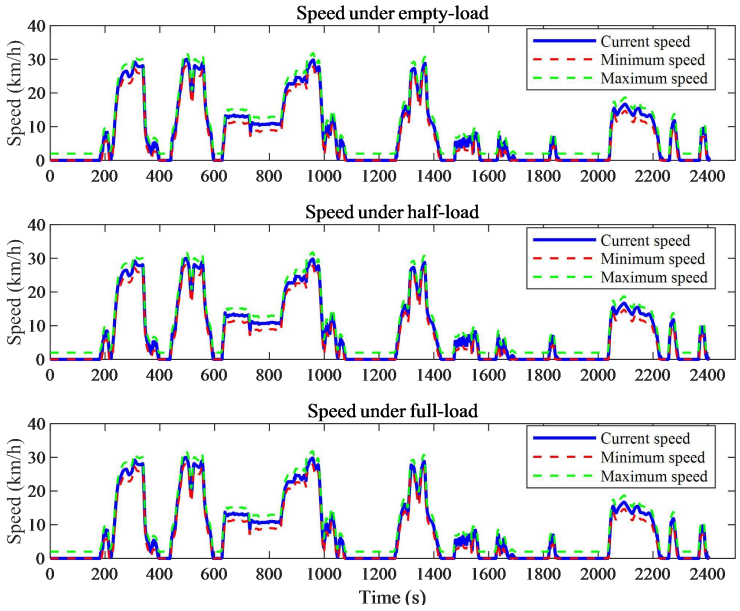

Figure 15. The real-time speed tracking situation of the tractor during operation

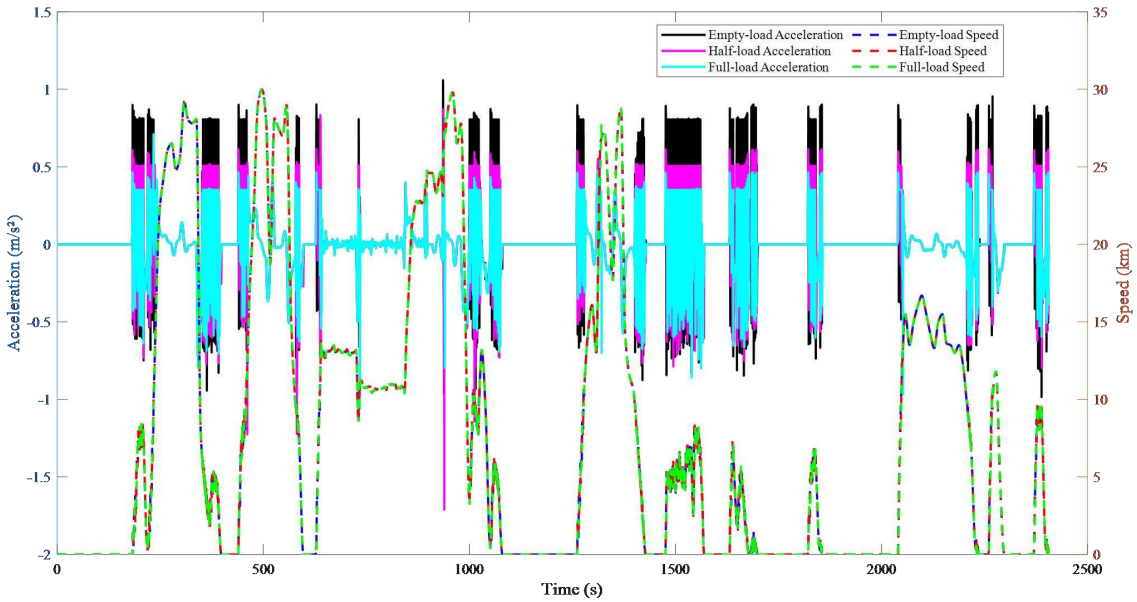

Figure 16. The acceleration response of tractors under different loads

Table 9. Comparison between simulation and test results of tractor fuel consumption

Load	Empty-load	Half-load	Full-load
Simulated fuel consumption ($L \cdot 100\ km^{-1}$)	21.65	27.09	33.30
Test fuel consumption ($L \cdot 100\ km^{-1}$)	23.22	29.31	36.84
Error (%)	6.76	7.57	9.61

5 Conclusion

In summary, this study proposes a measurement-oriented, data-driven framework for constructing representative transportation operating cycles of high-horsepower tractors by

integrating DBI-guided K-means clustering with VSP per mileage. The main findings can be summarized as follows:

(1) The synthesized operating cycle effectively captures the coupled speed-acceleration characteristics of tractor operation under mixed field-road conditions. Compared with conventional cycle-construction methods relying on empirically selected cluster numbers, the proposed framework provides a more objective and measurement-based representation of complex and highly variable tractor operating behaviours.

(2) The constructed cycle demonstrates high fidelity in both statistical distribution and temporal correlation characteristics. The average deviation of representative kinematic features relative to the full measured dataset is 5.28%, while the RMSE between the ACF curves of the measured samples and the synthesized cycle is only 0.0312. These results confirm that the proposed cycle preserves not only the global statistical characteristics but also the intrinsic short-term temporal continuity of real operating conditions.

(3) Cruise-based validation further verifies the practical applicability of the constructed cycle in energy-consumption analysis. Under empty-load, half-load, and full-load conditions, the discrepancies between simulated and measured fuel consumption remain within 10% (6.76%, 7.57%, and 9.61%, respectively), demonstrating that the synthesized cycle can effectively reproduce real-world operational behaviour within acceptable measurement uncertainty.

(4) The developed operating-cycle database provides a standardized measurement basis for the evaluation and comparison of tractor operating behaviour. It addresses the current gap in measurement-based transportation cycle research for high-horsepower tractors and supports further applications such as energy-consumption assessment, powertrain optimization, and control strategy development in agricultural machinery systems.

Data availability statement

The data will be available to access if needed.

Acknowledgements

The authors would like to appreciate the support from Jiangsu Modern Agricultural Machinery Science and Technology Demonstration Park. The work was supported by National Key Research and Development Program of China (grant number 2022YFD2001202B), Jiangsu Province Agricultural Science and Technology Independent Innovation Fund Project (grant number CX (24) 3028), Jiangsu Province Higher Education "Blue and Green Project" Funding Project, High-speed Train Key System Integration and Technology Innovation Centre (grant number KYPT2025001), Research on Key Systems of CR450 EMU Technological Innovation Project (CXTD2025002), Basic Science (Natural Science) Research Project of Higher Education Institutions in Jiangsu Province (22KJB580009).

References

[1] Xue Yuqiang, Jiang Rongchao, Zheng Xuguang. Construction Method of Commercial Vehicle Driving Conditions Based on Internet of Vehicles Data [J]. Highway Traffic Science and Technology, 2023, 40(8): 192-198. (in Chinese with English abstract)

[2] Lakhey O B, Bista A, Silwal B, et al. Significance a standard terrain-based drive-cycle for light electric vehicles (EV) performance evaluation[C]//2024 IEEE International Conference on Power System Technology (powercon). 2024: 1-5, https://doi.org/10.1109/PowerCon60995.2024.10870557.

[3] Hu Z, Wang Z, Lu Z, et al. Effect of idle-start-go on the economy and emissions of gasoline vehicles[J]. Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering, 2024, 238(1): 159-171, https://doi.org/10.1177/09544070221121868.

[4] KoteswaraRao K V, Srinivasulu G N, Rahul J R, et al. Optimal component sizing and performance of Fuel Cell−Battery powered vehicle over world harmonized and new european driving cycles[J]. Energy Conversion and Management, 2024, 300: 117992, https://doi.org/10.1016/j.enconman.2023.117992.

[5] Zhou Yingyu, He Ling, Chen Jiadui, et al. Based on UMAP-KNN method of bus operating mode to build [J/OL]. Mechanical design and manufacturing,1-7[2025-10-03], https://doi.org/10.19356/j.cnki.1001-3997.20250919.006.

[6] Song Yuzhen, Wu Zhimin, Yin Xiaofeng, et al. Research on Driving Conditions on Urban Slopes Based on Principal Component Analysis and Neural Network Clustering [J]. Automotive Technology, 2025(5): 47-54. (in Chinese with English abstract)

[7] Hu Julin, He Hongwen, Han Xuefeng. Construction Method of Multi-Road surface - three-dimensional Working Conditions for Tracked Vehicles Based on Micro-Motion Fragments [J] Journal of Armaments, 2020,46(09): 249-263. (in Chinese with English abstract)

[8] Hull C, Collett K A, McCulloch M D. Developing a representative driving cycle for paratransit that reflects measured data transients: case study in Stellenbosch, south Africa[J]. Transportation Research Part A: Policy and Practice, 2024, 181: 103987, https://doi.org/10.1016/j.tra.2024.103987.

[9] Almachi J C, Saguay J, Anrango E, et al. Clustering-Based Urban Driving Cycle Generation: A Data-Driven Approach for Traffic Analysis and Sustainable Mobility Applications in Ecuador[J]. Sustainability, 2025, 17(8): 3353, https://doi.org/10.3390/su17083353.

[10] Li Bin, Liu Hongli, Song Rui, et al. Construction of Driving Conditions and Analysis of Gear Shifting Scenarios for AMT Semi-trailer tractors [J]. China Journal of Highway and Transport, 2024, 37(9): 312-326. (in Chinese with English abstract)

[11] Zhang Haoran, Liu Yutao, Yan Xiaoxia, et al. Construction of Multi-dimensional Driving Conditions for Urban Rail Trams [J]. Journal of Electrical Engineering, 2024, 19(2): 353-363. (in Chinese with English abstract)

[12] Rueda D R. Development of AI-based typical driving cycles using variations of the k-means clustering algorithm[J]. 2023, http://hdl.handle.net/10754/689917.

[13] Zhang J, Wang Z, Liu P, et al. Driving cycles construction for electric vehicles considering road environment: A case study in Beijing[J]. Applied Energy, 2019, 253: 113514, https://doi.org/10.1016/j.apenergy.2019.113514.

[14] Bhatti A H U, Kazmi S A A, Tariq A, et al. Development and analysis of electric vehicle driving cycle for hilly urban areas[J]. Transportation Research Part D: Transport and Environment, 2021, 99: 103025, https://doi.org/10.1016/j.trd.2021.103025.

[15] Sithananthan M, Kumar R, Saxena D. Development of Indian motorcycle driving cycles, evaluation for fuel economy and emissions[J]. Environment, Development and Sustainability, 2024, 26(3): 7015-7053, https://doi.org/ 10.1007/s10668-023-02997-5.

[16] Ashtari A, Bibeau E, Shahidinejad S. Using large driving record samples and a stochastic approach for real-world driving cycle construction: Winnipeg driving cycle[J]. Transportation science, 2014, 48(2): 170-183,

https://doi.org/10.1287/trsc.1120.0447.
[17] Yang Y, Zhao X, Yuan X, et al. A novel heavy-duty truck driving cycle construction framework based on big data[J]. Transportation Research Part D: Transport and Environment, 2024, 127: 104077, https://doi.org/10.1016/j.trd.2024.104077.
[18] Brady J, O' Mahony M. Development of a driving cycle to evaluate the energy economy of electric vehicles in urban areas[J]. Applied energy, 2016, 177: 165-178, https://doi.org/10.1016/j.apenergy.2016.05.094.
[19] Bishop J D K, Axon C J. Using natural driving experiments and Markov chains to develop realistic driving cycles[J]. Transportation Research Part D: Transport and Environment, 2024, 137: 104507, https://doi.org/10.1016/j.trd.2024.104507.
[20] Pouresmaeili M A, Aghayan I, Taghizadeh S A. Development of Mashhad driving cycle for passenger car to model vehicle exhaust emissions calibrated using on-board measurements[J]. Sustainable cities and society, 2018, 36: 12-20, https://doi.org/10.1016/j.scs.2017.09.034.
[21] Angelucci L, Mattetti M. The development of reference working cycles for agricultural tractors[J]. Biosystems Engineering, 2024, 242: 29-37, https://doi.org/10.1016/j.biosystemseng.2024.04.004
[22] Qiu H, Cui S, Wang S, et al. A clustering-based optimization method for the driving cycle construction: a case study in fuzhou and putian, China[J]. IEEE Transactions on Intelligent Transportation Systems, 2022, 23(10): 18681-18694, https://doi.org/10.1109/TITS.2022.316 0275.
[23] Hajduk P, Ranta M, Farzam Far M, et al. Enhanced socially oriented mission-based driving cycles generation and simulation framework for light electric vehicles[J]. Humanities and Social Sciences Communications, 2025, 12(1): 1166, https://doi.org/10.1057/s41599-025-05220-0.
[24] Akhtar S, Smith H, Caulfield B, et al. Markov Chain and K-means Clustering Analyses for Constructing Electric Vehicle Drive Cycles for Dublin, Ireland[C]//Transport Research Arena Conference. Cham: Springer Nature Switzerland, 2024: 145-151, https://doi.org/10.10 07/978-3-031-89444-2_20.
[25] Ma J, Pan M, Guan W, et al. Economy Optimization by Multi-Strategy Improved Whale Optimization Algorithm Based on User Driving Cycle Construction for Hybrid Electric Vehicles[J]. Machines, 2025, 13(2): 158, https://doi.org/10.3390/machines13020158.
[26] Peng F, Zhang Y, Song G, et al. Evaluation of Real-World Fuel Consumption of Hybrid-Electric Passenger Car Based on Speed-Specific Vehicle Power Distributions[J]. Journal of Advanced Transportation, 2023, 2023(1): 9016510, https://doi.org/10.1155/2023/9016510.
[27] Xu D, Gao Z, Guo Y, et al. Study on fuel consumption and emission characteristics of China VI heavy duty vehicle based on vehicle specific power[C]//E3S Web of Conferences. EDP Sciences, 2021, 268: 01055, https://doi.org/10.1051/e3sconf/202126801055.
[28] Sun Y, Tao H, Stojanovic V. End-to-end multi-scale residual network with parallel attention mechanism for fault diagnosis under noise and small samples[J]. ISA transactions, 2025, 157: 419-433,https://doi.org/10.1016/j.isatra.2024.12.023.
[29] Sun Y, Tao H, Stojanovic V, A Generic Single-Source Domain Generalization Framework for Fault Diagnosis via Wavelet Packet Augmentation and Pseudo-Domain Generation. IEEE Internet of Things Journal, 2025, 12(15): 31629-31642, https://10.1109/JIOT.2025.3573752.
[30] Tao H, Huang Z, Wang Y, et al. Efficient feature fusion network for small objects detection of traffic signs based on cross-dimensional and dual-domain information[J]. Measurement Science and Technology, 2025, 36(3): 035004, https://doi.org/10.1088/1361-6501/adb2ad.
[31] Wang Wenqi, Gao Guangchao, Zou Qingjian, et al. Construction Method for Vehicle Operating Conditions [J]. Journal of Highway and Transportation Technology, 2020, 37(09): 128-138. (in Chinese with English abstract)
[32] Jin Sihan, Peng Yiqiang, Wu Xiaohua, et al. Construction of Driving Conditions for Demonstration Operation of Fuel Cell Buses in Chengdu [J]. Journal of Automotive Safety and Energy Conservation, 2022, 13(1): 202-208.

(in Chinese with English abstract)

[33] Xu Tingting, Long Fangjia, Hu Xiaorui, et al. Operating Conditions of Pure Electric Vehicles in Mountainous Environments Driven by Big Data [J]. Highway Traffic Science and Technology, 2024, 41(1): 177-185. (in Chinese with English abstract)

[34] Ren H, Xu M, Lin T, et al. Working cycle condition construction for electric wheel loader based on principal component analysis and cluster analysis[J]. Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering, 2024, 238(10-11): 2905-2919, https://doi.org/10.1177/09544070241233627.

[35] Gurusamy A, Bokdia A, Kumar H, et al. Expediency analysis of clustering algorithms for electric two-wheeler driving cycle development under Indian smart city driving conditions[J]. IEEE Access, 2024, 12: 180279-180300, https://doi.org/10.1109/ACCESS.2024.3508754.

[36] Qin F, Lin J, Zhao Y, et al. A Markov Chain Based Method for Driving Cycle Segment Prediction of HEVs[C]//2024 4th Power System and Green Energy Conference (PSGEC). 2024: 519-523, https://doi.org/10.1109/PSGEC62376.2024.10720959.

[37] Rousseeuw, P.J., 1987. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. Journal of Computational and Applied Mathematics, 20: 53-65, https://doi.org/10.1016/ 037 7-0427(87)90125-7

[38] Wang Hongtu, Yu Lei, Hao Yanzhao, et al. Research on the Calculation Method of Specific Power of Heavy Motor Vehicles [J]. Safety and Environmental Engineering, 2011, 18(2): 124-128. (in Chinese with English abstract)

[39] Box, G.E.P., Jenkins, G.M., Reinsel, G.C., Ljung, G.M., 2015. Time Series Analysis: Forecasting and Control, 5th ed. Wiley, Hoboken, NJ, USA.